DESIGNING A CITY DATA POLICY
A Reference Guide

Contents

I. Executive Summary

With the rise of Smart Cities in India, many cities are embracing a digital culture based on the foundations of Information and Communication Technology (ICT) systems. Besides the government departments and agencies; the private sector organisations, corporates, community organisations, research and academic institutions are also playing a large role in the functioning of cities and implementation of such initiatives. These digital initiatives generate a huge amount of data across the country. This raw data can be harnessed to its full potential if utilised in a strategic manner. Appreciating the value of such data for improving efficiency and service delivery, the Ministry of Housing and Urban Affairs (MoHUA) is working on numerous data initiatives including the DataSmart Cities (DSC) Strategy, creation of a Data Maturity Assessment Framework (DMAF), among many others.

Cities being the engines of growth for the economy, can lead the path towards data-driven governance and evidence-based policy formulation if they are able to unlock the true potential of data. They need a robust mechanism to improve their overall data maturity and enhance their ability to handle the data. Thus, there is a need for each city to form a City Data Policy (CDP), a document that will lay down the principles on how data should be managed.

Such a policy will help the city in defining mechanisms around data governance, protection, collaboration and innovation. A CDP will help cities establish data standards and interoperable systems and enable efficient sharing of data among data owners across inter-and-intra governmental agencies. A CDP will provide an enabling provision and platform for proactive and open access to the data available with various organisations of the governments as well as other stakeholders relevant to city functioning. The importance of a CDP in the data ecosystem has been highlighted in the DSC Strategy and the DMAF.

Though some cities have come up with their city data policies, many of the cities are yet to do so. They need a proper framework to promote effective data sharing and exchange within the bigger ecosystem (Government, Industry, Academia and Communities), while at the same time managing the issues of data security and privacy. Consequently, the Ministry felt a need to develop a common governance framework containing guidelines and standards that define how to collect, store and manage data for cities.

This reference guide has been developed to facilitate the cities in designing their own CDPs. It essentially captures the need for forming a data policy at the city level and outlines a broad structure and key themes for a city to be able to manage and best utilise the data it generates. It is intended to assist city officials in making the overall process more structured and streamlined. From understanding why the creation of such a policy is important and beneficial to identifying and reaching out to stakeholders, this guide attempts to answer various questions cities may have on this subject.

However, it should be clarified that this document has been designed as a guide and not as a prescriptive document that cities must follow. Cities are expected to study this document, work with city departments and other stakeholders to contextualise the approach towards their own data policy based on the city's vision, its goals and its strengths. A city may choose to include certain elements in its CDP that are not covered in this document; not only is this expected but it is encouraged. This document can be used to gain an understanding of the underlying processes, ensure a strong foundation as well as generate new ideas that cities can take forward. The document concludes with a checklist to further assist cities. Additionally, it includes a model CDP that cities can use as a template to draft their own CDPs.

II. Introduction

Recognizing the challenges and opportunities of urbanization in India, Ministry of Housing and Urban Development (MoHUA), Government of India (GoI), launched the Smart Cities Mission (SCM) in June 2015. With the launch of this mission, India started paving a new path towards transforming urban development using the power of digital technologies. Data being the fuel for these digital technologies, making cities 'DataSmart' was key in realizing the full potential of technology interventions and innovation ecosystem in cities. Thus, DataSmart Cities (DSC) Strategy document was released by MoHUA in 2019 that laid down the basic premise, foundational pillars and a suggested roadmap for cities to improve their readiness for intelligent use of data in addressing complex urban challenges.

DSC Strategy also lays down the tenets of a Data Maturity Assessment Framework (DMAF) to accelerate the establishment of data culture in cities. DMAF is intended to help city governments assess their data readiness on two tracks, Systemic and Sectoral maturity. The assessment aims to build a healthy spirit between cities by conducting periodic assessments. The DSC Strategy and the DMAF, focus on several activities such as building of governance structures, policy enablement, ensuring datasets compliance, maintaining quality with suggested guidelines, fostering data alliances, creating a CDP and conducting events such as hackathons, etc.

An important component of the DSC Strategy and the DMAF is the creation of a CDP. To unlock the power of data in the context of privacy, security and ownership in the context of the city, it is critical that cities create data policies that balance privacy, legal aspects and public benefit considerations. At the same time, the policy must define the contours of collaboration between various Governmental as well as non-Governmental entities on data sharing and access. Lack of a clear data policy prevents cities from adopting data-driven decision making due to underlying issues such as managing different types of data, data ownership and privacy.

Developing a comprehensive CDP provides clarity on the enterprise processes required to handle data and manage it on open data and exchange portal, helps in identification and segregation of datasets, and introduces accountability by nominating and engaging with departmental data officers. A data policy is also essential to understand the contours of data sharing, standardization, privacy, security and ownership in the context of the city. A CDP is critical to retain the trust between the city administration and its stakeholders, so stakeholders understand that the city data collection exercise is to their long-term benefit.

Thus, a well-thought-out CDP enables municipal departments and parastatal agencies to collect, aggregate and disseminate their data in a manner which enables inter- and intra-departmental data sharing, improves collaboration between government and external stakeholders, and assists in facilitating improved service delivery and increased socio-economic development in the city.

The CDP finds its basis in the already existing data policies and frameworks set out previously to collect and manage data. In the next section we will elaborate on some of these policies.

III. Need for Data Policy

Recognising the importance of having data that can be managed and shared easily among agencies, the GoI launched the National Data Sharing and Accessibility Policy (NDSAP) in 2012. The policy aims to facilitate access to Government-owned shareable data and information in both human-readable and machine-readable forms. The NDSAP applies to all shareable, non-sensitive data available either in digital or analog forms but generated using public funds by various Ministries, Departments, subordinate offices, organisations, and agencies of GoI, as well as of the States. It is designed to promote data sharing and enable access to public data for planning, development and awareness. Key principles on which data sharing and accessibility need to be based include openness, flexibility, transparency, quality, security and efficiency.

NDSAP defines standards and templates for publishing datasets and feeds on data portals to be adhered by the data officers. For instance, it mentions that all datasets are open by default unless classified as internal, sensitive, protected or restricted. NDSAP requires that all datasets and apps must be published along with proper metadata. Besides facilitating easy access to datasets, using a common data taxonomy/structure is extremely useful in the future for federation/integration of data catalogues. The policy recommends that datasets should be published in an open format, which can be accessed without the need for a software licence and should be machine-readable. It lists the various formats in which the data could be published.

The Open Government Data (OGD) Platform India **(https://data.gov.in/)** was launched in 2012 based on the NDSAP. OGD Platform has been set up to provide collated access to resources (datasets/apps) under catalogues published by different government entities in open format. The current version launched in December 2014 is the latest stable release of the platform. After the launch of the Digital India programme in 2015, OGD Platform has been included as one of the most important initiatives under "Information for All" – Pillar 6 of the programme. Further, the data usage licence should be used for datasets published under NDSAP and through the OGD Platform. The Cities could adopt the Government Open Data Licence (GODL), which can be accessed at: **https://data.gov.in/government-open-data-licence-india**.

Further to this, MoHUA's DSC Strategy emphasizes on a data governance framework, data categorisation and classification, roles and responsibilities of key stakeholders, formation of alliances etc. applicable to smart cities. While NDSAP is applicable to all entities of government setup, DSC Strategy is geared towards the Indian Smart Cities, and both define the basis for collecting and managing the city data.

As an extension to the OGD, the Smart Cities Mission launched its own Open Data Portal (**https://smartcities.data.gov.in/**) in February 2019. It is designed to be a single point of access to open datasets available with various Ministries/Departments in Smart Cities along with enhanced visualization, efficient discoverability of resources, widgets to share filtered set of data catalogues, and community participation through forums, blogs, infographics, and much more. The datasets here follow the principles laid down

in both NDSAP and the DSC Strategy. The portal is gaining traction with all 100 Smart Cities' City Data Officers (CDOs) already uploading datasets on the portal for their respective cities.

More recently, drawing inspiration from the General Data Protection Regulation, the Personal Data Protection Bill 2019 was tabled in the Indian Parliament by the Minister of Electronics and Information Technology in December 2019 which is expected to form the base of India's data protection architecture. The Bill aims to provide for protection of the privacy of individuals relating to their personal data. It seeks to govern the processing of personal data including collecting, recording, adapting, indexing, or even disclosing. These principles need to be kept in mind while handling of all personal data at the city level.

Although there are many national policies for management of data, there is a growing understanding that city governments' proximity to their constituents enables them to respond better to local needs. City governments can more efficiently understand, utilise and monitor the usage of data at the city level and put it to use for better service delivery. Thus, there arrives a need for a data policy at the city level, supporting data-driven decision making at the ground level.

Now that we have some idea about the various data policies already in place, in the next section we will be exploring some of the possible components that could be part of the CDP for a city.

IV. Components of a CDP

A CDP document will act as a guiding framework for the cities and hence should be properly defined and structured. It should provide guidance on various components of the data lifecycle that a city must follow. This section defines some of the components that a CDP must contain based on current frameworks and policies, including but not limited to, NDSAP, DSC Strategy and DMAF. Some of the possible components of the CDP may be as follows:

1. Data categorization

Different types of datasets generated by different Ministries/ Departments can be categorised into shareable data and non-shareable data. This applies to data generated both in geospatial and non-spatial forms. Non-spatial data is the type of data produced by a statistical system consisting of derived statistics like national accounts statistics, indicators like price index, census and surveys. The geospatial data, however, consists primarily of satellite data, maps, etc. In such a system, it becomes important to maintain standards with respect to metadata, data layout and data access policy. As per the DSC Strategy, any data can be categorized into two broad categories:

- **Personal data:** Personal data can be defined to have Personally Identifiable Information (PII). This would include data consisting of information which is related to an individual who can be identified from that information (or from that and other information in the possession of the data users). One of the most important requirements for data security is to ensure that personal data is not made public and that individuals cannot be identified through data. While personal data can be used by cities, this data must be anonymised before publishing.

- **Non-personal data:** Non-personal data is data without PII and includes anonymous data, namely information which does not relate to an identified or identifiable natural person. It is the personal data that has been rendered anonymous in such a manner that the data subject is no longer identifiable. Non-personal data includes personal data that has been anonymised and cannot be traced back to an individual, as well as more general data that does not relate to a person or persons such as traffic and transportation data, climate data, land use data, etc.

It should be noted that all data generated may be structured or unstructured data:
- **Structured Data:** Data which is generated by systems or humans and could be handled using existing or predefined models. Structured data could be stored in relational databases and analysed using basic search algorithms. E.g.: Location data, User transactions, Sensor data etc.

- **Unstructured Data:** Data which is generated by systems or humans and cannot be handled using existing or predefined models. Unstructured data cannot be stored in relational databases and is difficult to analyse using basic search algorithms. E.g.: Images, Video files, Audio Files etc.

2. Data classification

Data classification is the process of organising data into categories for its most effective and efficient use. Further to data categorization, there is a need to classify the data basis its intended usage and stakeholders. There may be some data which can be open to the public, while some other may be confidential and restricted. Such distinctions need to be appropriately defined to prevent misuse and maintain confidentiality.

Classification	Class	Definition
Level 1	Public/Shareable Data	Those data not covered under the scope of negative list and non-sensitive in nature. This data is available for public consumption and use.
Level 2	Negative List	Non-shareable data as declared by the departments/organisations.
Level 3	Restricted Data	Data which are accessible only through a prescribed process of registration and authorization by respective departments/organisations.
Level 4	Sensitive data	Sensitive data as defined in various Acts and rules of the Government of India.

The DSC Strategy further goes to classify the data to highlight whether the data is protected or only for internal use. For each of the dataset being stored within the city, proper tagging should be done to clearly define what is the level of shareability of that particular dataset, such as if it is open data or can be shared with a licence. This will eventually help as a quick reference check when data needs to be shared with other parties.

## 3. Data flow/approval framework

Once the data is collected, various stakeholders may be involved in its movement from source to destination. This is true for any process including getting data from on-field surveys, uploading datasets on the smart cities portal or even publishing blogs or visualizations on the portal. This movement needs to be adequately monitored to avoid any undue changes and maintain data authenticity and correctness.

City administration needs to establish specific processes to manage the flow of available data through the various channels. There needs to be a formal framework defining the movement of data among the various stakeholders. At every stage of data management - generation, cleaning, aggregation, analysis, dissemination etc. - the concerned stakeholders need to approve and authorise it. The data flow could be organised through regular reports obtained from various data coordinators. Data flow for electronic data shall maintain a timestamping of each modification/review and the movement needs to consider a versioning system of files.

The approval process should be followed on by the officers nominated/appointed for this purpose. The accountability of the nodal officer should be fixed. The process may involve analysis and verification of data as per the type of data structures defined in the CDP. Data along with the metadata contributed by the Data Coordinators should pass through a predefined workflow to the CDO; who in turn should

ensure that it complies with the CDP and other State level and National Level Data Policies. Data may be published only after approval of CDO or Municipal Commissioner as designated earlier.

*Some of the cities recognise the importance of having in place a clearly defined data flow framework in their CDP. A sample of the framework for uploading data on open data portal defined by **Pimpri Chinchwad** in its CDP is given below:*
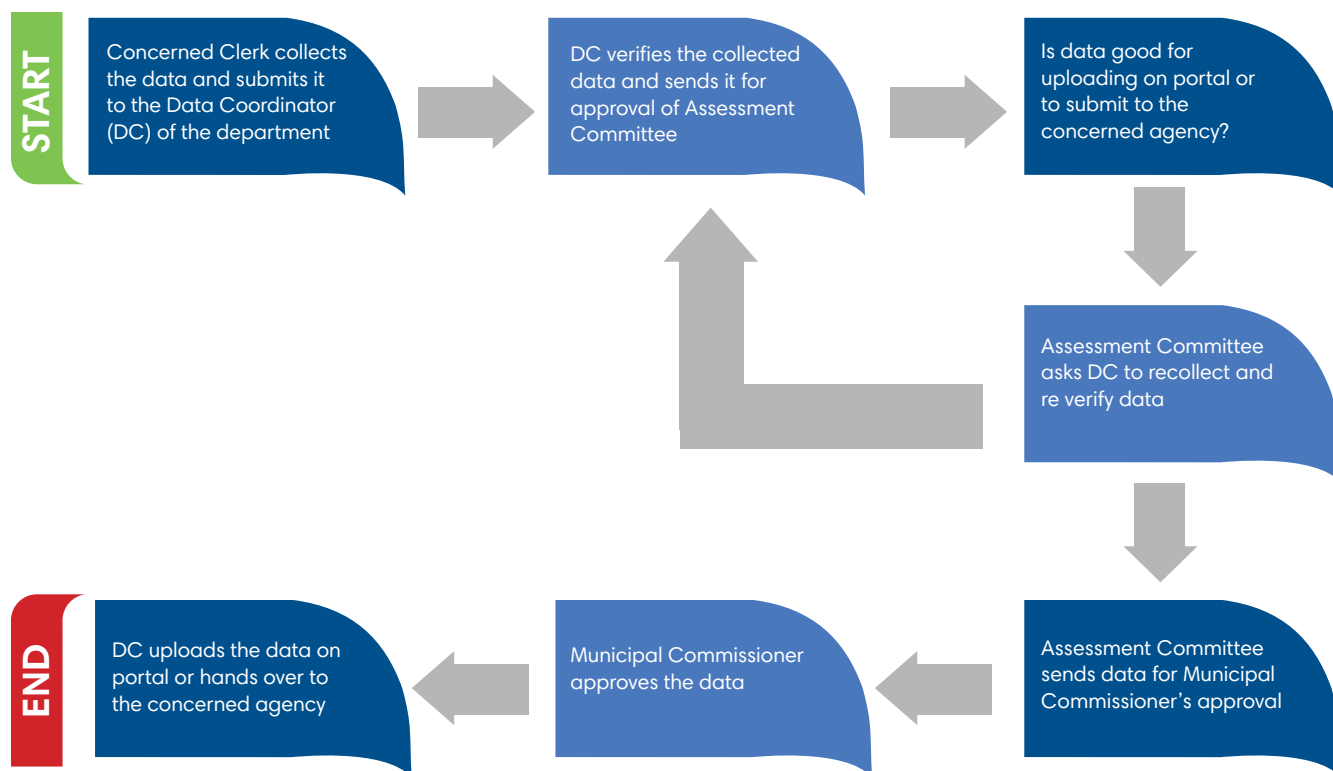
4. Data archival and retention

Data Retention

Retention is defined as the maintenance and upkeep of data which can be accessed by an authorised user as and when required at a later stage. Based on certain aspects such as the need for documents to resolve queries, inter-department requests and responses to on-going audit requirements, the CDP should define data retention policies to be followed. Retrieval schedule of the data will be as per the rules and regulations defined for the same.

It is necessary for cities to be clear about their data retention strategies. This includes what data should be retained; where data should be retained; the way in which data should be retained, including backups and categorization of data; how long data should be retained; and the processes of data retrieval.

Data Archival

Data archiving may be defined as a collection of historical records specifically selected for long-term retention and future reference. Data archives consist of previously collected data that is still important and necessary for future reference, as well as data that must be retained for regulatory compliance. A time

frame should be defined until which a data remains in the retention stage and after which it is moved to the archive stage. Few things to be kept in mind:

- Ideally, all city data should be archived in a single location and the archive should be backed up
- When datasets are defined and designed, archived data must also be categorized, catalogued, and should be easily searchable
- Access restrictions need to be put, in terms of authorisations
- It needs to be ensured that the data is protected and cannot be tampered with

Solapur's CDP defines that "data will be stored in the main database for 6 Months in a live state so that whenever a report needs to be generated, the data will be extracted from the main database. Data older than 6 months will be archived. If report duration extends beyond 6 months, the data will be retrieved from archival to generate the report."

5. Data security and privacy

Managing security and privacy of data is crucial to building and maintaining trust among all stakeholders and is a critical element of the CDP. Data collection, sharing and analysis must be ring-fenced by a privacy first approach to guarantee protection for citizens. Cities should develop ethical frameworks for data ownership and privacy which overcome any gaps in current legislation. CDP should also talk about the specific guidelines for handling the personally identifiable information in order to safeguard interests of the data owners, including seeking consent from owner, and giving a disclaimer on the use of the data collected. Personal data collection must be supported by pre-defined purposes that justify the collection of citizen's personal data. The specific purpose of collected data used or shared by the municipality should also be documented.

Data security works at different levels. On the one hand it must ensure that restricted data has proper authorization checks in place, on the other it must ensure that data is not corrupted. This also means that data should be properly backed-up and stored like it has been discussed above. The CDP should clearly define procedures for various security areas including, but not limited to, system security, network security, data protection, citizen privacy, application security, and database monitoring. Outlining protocols and mechanisms for grievance redressal in case of a data breach or the failure to protect data privacy/ restrictions is critical. It should also be ensured that the CDP adheres to the existing National and State level data policies on security safeguards.

Several CDPs have tackled the issue of data privacy and security in their own manner. Bhopal CDP, for example, recommends that "all data access must be through Application Programming Interface (API) calls to ensure appropriate security controls. Any kind of data sharing should comply with existing standards and certifications for data privacy and security. For open data, it is recommended that direct access to data be prohibited and use of APIs mandated. Data dissemination should be only to authenticated and authorised stakeholders (both internal and external) through data fiduciaries."

Studying these components will give a fair idea on how the data needs to be maintained. We will now look at some of the standard operating procedures that should be followed in specific circumstances.

STANDARD OPERATING PROCEDURES (SOPs)

Standard Operating Procedures (SOPs) are succinct guidelines designed to achieve consistency in specified situations by postulating a standard practice in performing those functions. These SOPs are designed with a view to enhance and standardise data operation and management.

6.a) SOP for data collection

Data collection is referred to as the method of collecting information in a systematic way. The process comprises of solutions to customize and prepare the data collected through various sources and formats i.e. structured and unstructured for analysis purpose. Thus, there should be predefined methods for the collection and management of such data held on behalf of the city.

CDP needs to define the various sources from where the data should be collected. These data source may include:

- **Field Data:** Smart Cities are deploying various Internet of Things (IoT) sensors, devices, cameras and solutions to capture the data from the field directly.
- **Operations Data:** Multiple systems are deployed under various departments to manage city core operations like Water Supply, Surveillance, Traffic Electricity, Street Lights, Water and Sewerage Treatment, Health, Education, Fire Department, Disaster Response and Licence permits etc. These systems generate various data points which provides vital information and intelligence to officers to manage critical services of the city.
- **Third Party Platform and Mobile Apps:** Citizens also avail services from private businesses like radio taxi, food delivery, hospitals and labs etc. which could provide actionable insights and patterns that may be instrumental in policy formulation and city planning.
- **Internet:** Various platforms which capture general sentiments of citizens and communities may provide insights on their opinions towards specific issues.

The SOP may also consider:

- incorporating consent before/at the time of collection of any personal data from citizens, with higher standards of consent being applicable to sensitive Data
- incorporating clear provisions about data processing including the identity of the entities that are processing the data, the purpose for which it is being processed, the entities that may access this data among other details.

6.b) SOP for electronic data collection

Electronic form means maintenance of documents in any electronic device such as computer, laptop, compact disc, space on a cloud platform or any other form of storage and retrieval device, considered feasible, whether the same is in possession or control of any agency.

Sources from which electronic data needs to be collected should be evaluated, i.e. Electronic Sensors, IT systems or any other source. While collecting such data, the IT applications/IT systems should be developed in such a way that they follow proper checks and conditions, and under any circumstances do not accept garbage/wrong data/duplicate data/null data. The SOP should evaluate and define the sources of data collection, frequency of data, storage methods, and the checks that should be in place to maintain the accuracy and privacy of such data.

It should be noted that ease of data collection should be a priority at data collection points since the users collecting electronic data may not be technological sound in some cases. At times the resources deployed at the last mile may not understand the value of meticulous and focused data collection and may need to be trained not only on the tools but also on the importance of collecting data properly.

7. SOP for data processing and cleaning

To make data more meaningful and useful, it must be processed. Before data is processed, the dataset must be cleaned since the data may not have been captured in the same format. This is an important step that can ensure the accuracy of results post any data analysis. This SOP should specify the formats, process flow, responsibilities of key people involved in data cleaning and processing.

Data cleaning includes checking for missing variables, eliminating redundancy, removing or rectifying incorrect or inaccurate data, ensuring uniformity in measures, etc. Errors or inconsistencies in the dataset will inevitably be reflected in analysed data. It is the responsibility of the Data Coordinators to ensure that data is collected in a normalized format and follows all compliance rules such as for blank cells, special characters, spaces, etc.

While governments should make public datasets accessible to citizens in their raw format (citizens might want to utilise these for their own research and analysis), the government may also present some processed data. This could be in the form of tables, graphs, charts, infographics, or descriptions and accessible through APIs wherever possible. The most appropriate way of presenting data will differ according to the datasets and the information being conveyed. This will help in drawing insights from the raw data and enable outcome-based analysis.

Nagpur has provided proper workflow details of the various SOPs. Workflow details of SOP for data processing and cleaning is highlighted below:
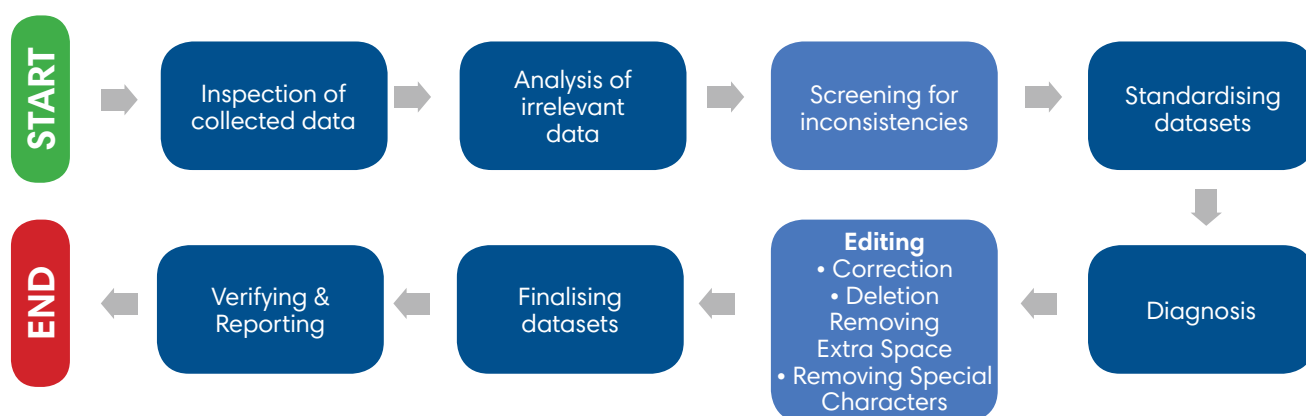
8. SOP for quality assessment of datasets

The data thus obtained needs to be further processed to ensure that the quality standards are met. The data should meet the norms of quality which may include but are not limited to, accuracy, precision or resolution, completeness, format, consistency, range, pattern etc. The best practices in quality assessment and datasets management should be clearly laid down. The responsibility of relevancy and quality of datasets published rests with the CDO.

Data quality cannot be ensured through data release alone; efforts need to be made to keep the data up-to-date, accurate and accessible. Data with serious accuracy and quality concerns should be adequately documented to avoid creating confusion or spreading misinformation. Each update should include clear and complete metadata (including a conspicuous contact person), group datasets where appropriate, and address concerns noted via a prominent feedback mechanism.

9. SOP for data publishing

Data should be published as per the open data norms and in accordance with the NDSAP. The selection of datasets and the data publishing options should be carefully studied before publishing the data on any of the portals. When the data is published it should be tagged with the appropriate metadata. Automated flow of data i.e. data by default mechanism should be given prominent focus to publish more high value datasets in the public domain. These processes of data flow should be defined without any human intervention. Machine-to-machine technologies such as APIs, Web Services, data from IoT devices, etc. should be prioritised.

For published datasets, a proper mechanism for user feedback and grievance redressal should be defined. The feedback could be on data accuracy, quality, frequency, granularity and other related issues faced by the data users. This will help the data providers to ensure most relevant datasets are published on the portal.

It is necessary for the city to determine the frequency with which each dataset that it publishes will be updated and where archived datasets can be found. For datasets that are published routinely, governments must clarify a change in the definition of parameter of the variable. City governments should also clarify whether datasets are only published digitally, the formats in which they are published digitally, and whether hard copies of the datasets can also be found. Care should be taken in keeping the concerned stakeholders updated about the new guidelines and open data norms.

While uploading datasets/apps forms the major activity of implementing the CDP, it is equally important to continuously maintain these datasets to ensure they remain relevant and valid. The CDO must ensure that the datasets are updated in case if any changes/amendments occur in the source of data i.e. the primary data. It should also be ensured that errors and inconsistencies are removed from the primary data and that the dataset included all the metadata along with referential information. We may also have the datasets directly linked to the APIs, so that any changes in the base data are simultaneously reflected in the portals. Wherever possible, real-time information updates are recommended as they would maximize the utility the public can obtain from the information. For the rest, it will help to have a regular maintenance schedule in place and ensuring compliance with the same.

10. SOP for engaging stakeholders

There should be a well-defined process for engaging stakeholders to assess the data needs of the City. The City should recognize the value of engaging all four stakeholders of the quadruple-helix model – Government, Industry, Academia and Citizens for this aspect. CDOs along with the team of Data Champions/Coordinators must assess the data requirements of various stakeholders in smart city

ecosystem. External as well as internal stakeholders need to be engaged at an operational, tactical and strategic level to assess the needs. Data needs and frequency of consumption should also be outlined. The SOP shall also define the rights and duties of the stakeholders. The following partnership stages detail the city's approach in building partnerships:

1. **Having an internal kick-off within department staff and other urban focused institutions:** CDO may identify three to four critical areas for the city. He will need to examine the system to identify the real issues city is trying to solve and to understand the relationships in the system.

2. **Scanning the ecosystem for partners:** Based on the dimensions of domain knowledge, influence, public outreach, financial leverage, mandate and incentives, CDO may identify the types of partnerships required.

3. **Having an external and invitational kick-off with partners:** CDO should have a roadmap on the critical areas for data collaboration including mission, vision, time, cost, risks and scope.

4. **Operationalising the partnerships:** CDO should ensure effective implementation of the partnership with appropriate documentation, key performance indicators and deliverables. CDO needs to mobilise action from different stakeholders with different roles and power, co-define the process, co-develop solutions, and co-deliver actions and adopt an incremental, problem-driven, iterative approach that promotes experimentation, innovation and learning.

5. **Larger public outreach:** CDO should enable outreach via communities such as co-working spaces, developer groups through mechanisms such as data meets, hackathons, developer communities etc.

11. SOP for data collection, processing and analysis for on-field surveys

Depending on the requirement of data, competent agencies can be employed to conduct field surveys. Periodic gathering and collection of datasets as per defined frequency and granularity may be done. This will help in collection of primary data e.g. Population Census, Education Census, Economic Survey, etc. Surveys involve administering a questionnaire to the target group of people across the sample area. The observations, responses and recorded facts then need to be integrated and organised in a systematic and logical framework to be available for further analysis. The SOP should include the methods in which data should be collected and further processed from on-field surveys.

Data collected from the field is usually in raw form of statements, facts and qualitative terms. This data may contain errors, omissions and inconsistencies, and require further processing to be deemed useful for analysis. The survey responses should be verified through cross checks and mapping with the on-ground situation.

With time, the survey process should transition from paper-based manual one to automated processes without much intervention of humans. As most of the process would be automated and handled by an e-mode, data will be available more easily available for further analysis.

12. SOP for data analysis

Apart from presenting the actual data, cleaned datasets might also be used to discover new patterns or analyse existing patterns, trends or behaviours. Data analysis may help multi-disciplinary researchers provide different perspectives or even solutions on civic issues like transport, traffic, solid waste etc. In line with this, the SOP should define the contours of analysing the collated data. It should shed light on how the data could be utilised to provide meaningful analysis that can help with solving urban problems.

Data analysis comprises of tools and methods used to process structured and unstructured data on various dimensions for various purposes. This will help users derive information, intelligence and knowledge out of processed data. Further emerging technologies such as analytics, artificial intelligence, machine learning etc. may be used for generating insights from the data gathered from the multitude of sources which will help in improved decision making. For example:

- Trend analysis and pattern identification on time series data (days, weeks, months, quarters or seasonal etc.)
- Trend analysis and pattern identification using various dimensions: cost, budget, domain-specific parameters etc.
- Comparison between various parameters in different geographies etc.
- Visualization to view the trends and patterns for decision making, by converting the data into a more comprehensible and user-friendly format.

Pune has also placed a great focus on defining the responsibilities of the key stakeholders. Pune has taken help of RACI Matrix, a powerful tool to assist in the identification of roles and assigning of cross-functional responsibilities.

RACI Definitions:

Responsibility = person or role responsible for ensuring that the activity is completed

Accountable = person or role responsible for actually doing or completing the activity

Consulted = person or role whose subject matter expertise is required in order to complete the activity

Informed = person or role that needs to be kept informed of the status of activity completion

Sample from Pune SOP for Electronic Data Collection is highlighted below:

Stakeholders Open Data Activities	Municipal Commissioner	Chief Data Officer	City Data Officer	Data Champion	Data Coordinators	Other Representatives (Alliance/ Dept. SPOC etc.)
Identification of data sets	I	I	A	C	R	I/C
Defining standard format or data set format for collection/ gathering of data	I	I	A	C	R	I/C
Data collection	I	I	R	C	A	C
Data sanitization	I	I	A	C	R	I/C
Data categorization and classification	I	I	A	R/C	R	I/C
Data Standardization	I	I	A	R/C	R	I/C
Data Approval	I	A	R	C	I	I
Publishing the data sets	I/C	C	A	I	R	I

V. Policy Implementation and Governance

Creating the policy solves one part of the puzzle; effective implementation is the rest. The City needs to ensure the CDP is implemented as per the guidelines laid down in it. A central mechanism should also be defined to monitor the implementation and other related activities of the policy at the city level at regular intervals. There need to be defined roles and responsibilities for the key stakeholders and those driving the policy at the city level. Also, there needs to be cohesiveness among such frameworks being implemented at the national and state levels. One may refer the DSC Strategy for more details. Below are some key aspects that should be kept in mind while designing the CDP:

DATA GOVERNANCE

Only having the data is not enough, it should be reliable, trusted and in a form, which can be utilised further. This leads to the need of a Data Governance Framework, which will define the decision-making process and authority for data-related matters. Incorporating accountability measures such as reporting requirements will strengthen data governance within the city. These may include creating public registries of all data sharing agreements and data exchanges that the city may have undertaken, or periodic reports on the expansion of data collection points, the infrastructure being used to collect data, etc. A robust CDP should also ensure compliance with National and State level policies such as National Data Sharing and Accessibility Policy (NDSAP), and the Personal Data Protection Bill (2019).

ROLES AND RESPONSIBILITIES

The CDP should also define the roles and responsibilities of the key people including the City Data Officer, Data Champions and Data Coordinators. They should work with city leadership to assess and tap the potential of data, establish a data culture across the organisations, help with the implementation of the CDP in their respective departments/organisations, aggregating data demand, and ensuring data quality. This will help in promoting accountability and transparency.

DRIVING DATA COLLABORATION AT STATE AND NATIONAL LEVEL

Though the CDP may be written for a specific city, there needs to be a close collaboration and integration with State and National Level bodies. The entities involved at various levels should work together seamlessly and adhere to set protocols and guidelines as prescribed by the GoI. A City Data Cell may be formulated to manage the data platforms, provide technical advice to the departments, handhold for dataset contribution as well as capacity building of Data Coordinators and CDOs.

CITY DATA ALLIANCE (CDA)

The CDA will provide a collaborative framework to create and define use cases to solve critical city problems using data and undertake continuous dialogue between various stakeholders in the city around the CDP to inform and evolve it effectively. Key stakeholders should come together to set up a City Data Alliance to assess, strategize, plan, implement and review the CDP, including government agencies, industry, academia, citizens and communities, etc. This will help in keeping the CDP relevant and updated.

Jabalpur and Bhopal have defined a comprehensive City Level Institutional Governance Structure to ensure creation and effective implementation of their CDP. The primary goal of any institutional governance mechanism is to deliver a set of rules which are derived from the principles of data governance, i.e., integrity, transparency, auditability, accountability, stewardship, checks and balances, standardization, and change management. Their City Level Institutional Governance Structure sheds light on various actors and their core responsibilities at the city level.

1	2	3	4	5
City Data Officer	**Data Champions**	**Data Coordinators**	**City Data Cell**	**Smart City Data Alliance**

VI. Conclusion

Data and digital technologies have helped streamline the approach to development of the urban ecosystem. More profoundly, data and digital technologies can be used to identify the most pressing issues in a city while also offering possible avenues to solve these problems. Data is at the crux of today's smart cities, however for it to be properly utilised, it needs to be adequately understood, collected and disseminated to various stakeholders. Since the current methods of collecting and storing data are fairly diverse, it is necessary to develop a comprehensive framework that promotes a specific method of storing and sharing data to enable all users to access and use it.
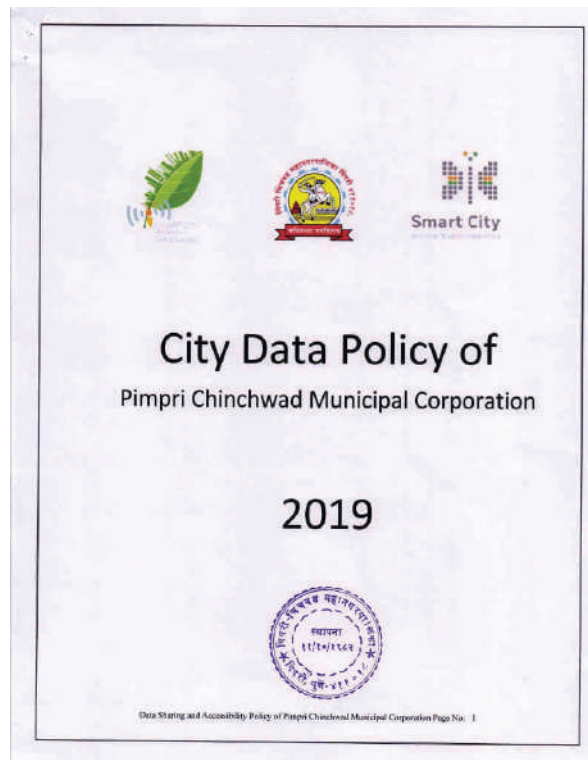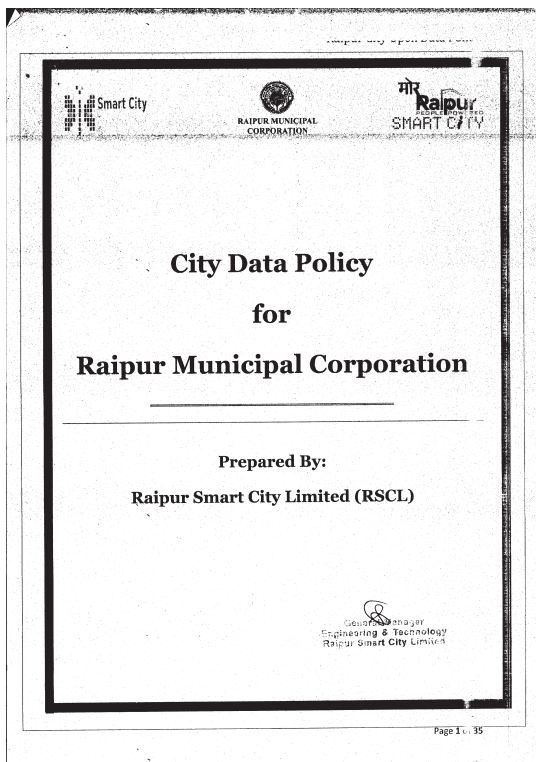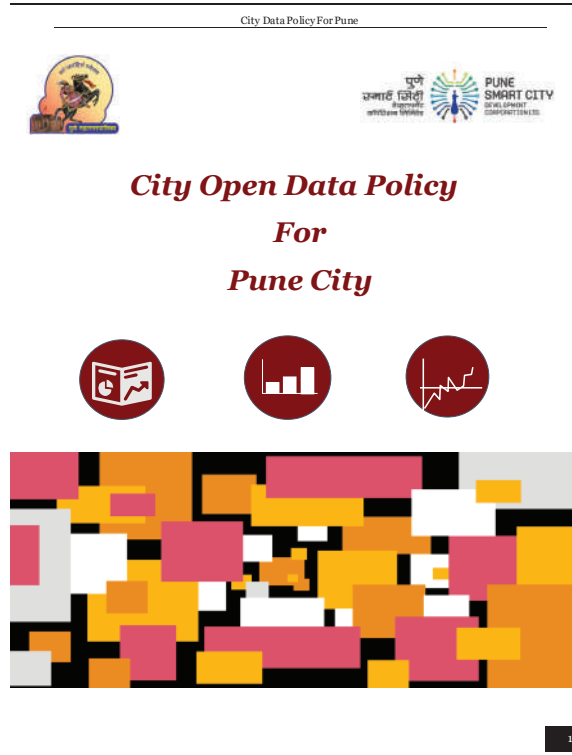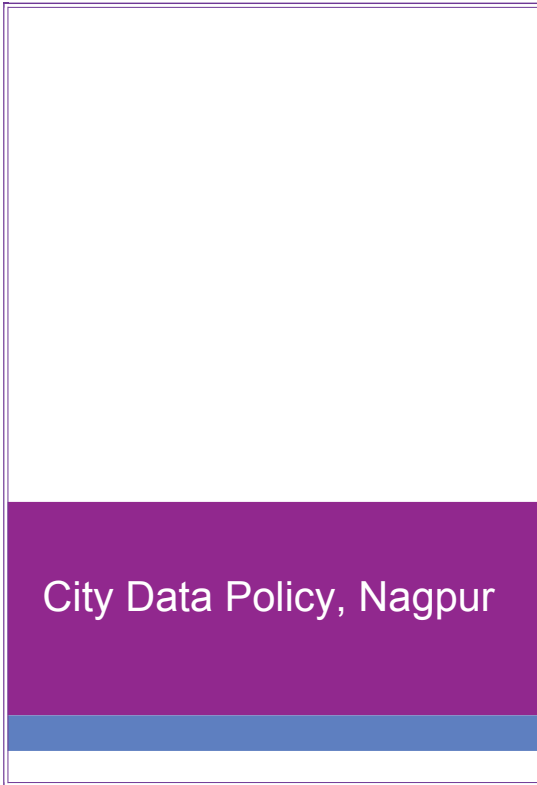
A CDP document aims to promote a technology-based culture of data management as well as data sharing and access. It is a carefully thought of system of norms to guide decisions and achieve rational outcomes from the data available in the city. It proactively opens up the data, which could be further utilised for developmental purposes and solving urban challenges. Cities should continue the process of evolving the policy further, keeping in tune with the technological advancements and the National, State and City requirements.

The aim of this guide is to help cities formulate their own data policies. Similar to the DSC Strategy and DMAF, this guide needs to be considered as an indicative reference. Cities may choose to concentrate on and incorporate select parts of this guide and may even add on new elements as they see fit. It should be noted that while CDP will provide the official mandate, proper facilitation for accessibility and usability of data by the implementers will provide added value to all stakeholders.

- **Data:** Data refers to a representation of information, numerical compilations and observations, documents, facts, maps, images, charts, tables and figures, concepts in digital and/or analog form collected together for reference or analysis.

- **Data Archive:** A place where machine-readable data are acquired, manipulated, documented and distributed to others for further analysis and consumptions.

- **Data Generation:** Initial generation/collection of data or subsequent addition of data to the same specification.

- **Dataset:** A named collection of related sets of information composed of separate elements, but which can be manipulated as a unit.

- **Geospatial Data:** All data which is geographically referenced.

- **Information:** Processed data is referred to as Information.

- **Metadata:** Metadata is data about data. The information that describes the data source, and the time, place, and conditions under which the data were created. Metadata informs the users of who, when, what and where data were generated. Metadata allows the data to be traced to a known origin and known quality.

- **Negative List:** List of prohibitive datasets/feeds, deemed non-shareable by the departments/ organisations.

- **Restricted Data:** Data which are accessible only through a prescribed process of registrations and authorization by respective departments/organisation since it could lead to a threat to life or loss of public assets or critical infrastructure.

- **Shareable Data:** The data not covered under the scope of negative list and non-sensitive in nature falls under shareable data.

- **Standards:** Any application that embeds data handling functions (e.g. data collection, management, transfer, integration, publication etc.).

- **Open Access:** Access to data generated from public funding should be easy, timely, user-friendly and web-based without any process of registration/authorization.

City Data Policy, Nagpur

City Open Data Policy
For
Pune City

City Data Policy

for

Raipur Municipal Corporation

Prepared By:

Raipur Smart City Limited (RSCL)

City Data Policy of
Pimpri Chinchwad Municipal Corporation

2019

Annexure III: A CDP checklist for cities

✔ An overall framework for creating the city's data policy is developed.

✔ Scope of the policy is clearly laid out.

✔ Key stakeholders who will be involved in the process of creating, writing and reviewing the policy are identified.

✔ The specific components that would form a part of the CDP are identified.

✔ Information about each component is collected so that all aspects of data are appropriately covered.

✔ CDP adheres to the principles laid down under NDSAP, DSC Strategy and DMAF.

✔ CDP includes principles of data classification and categorization, data flow/approval framework, archival and retention policies, and tenets of data security and privacy.

✔ There is adequate information around formation of City Data Team and City Data Alliance

✔ There is a clear indication on the access policy of the datasets along with their classification.

✔ Ample space is available for data storage, keeping in mind the retention and archival timing.

✔ Approvers for various data workflows are defined and intimated.

✔ CDP includes pre-defined and agreed upon procedures for data collection, processing and cleaning of data, quality assessment and publishing, as well as engaging stakeholders.

✔ Responsibilities of various stakeholders are defined.

✔ The policy covers key aspects for its successful and effective implementation including a data governance framework, defining roles and responsibilities of key participants and monitoring policy implementation.

✔ CDP is reviewed by the identified stakeholders and their inputs have been incorporated.

✔ CDP is finalised and approved by the relevant authority for public circulation.

City Data Policy

for

<<City Name>>
MUNICIPAL CORPORATION

Disclaimer

This Model City Data Policy document has been prepared as guidance on creating a CDP for cities. The information contained in this document on behalf of the Government of India, Ministry of Housing and Urban Affairs (Smart Cities Division), is provided to cities to guide them to formulate their own City Data Policies.

Information provided in this Model CDP is on a wide range of topics related to data, some of which depend on the interpretation of law. This information given is not an exhaustive account of statutory requirements and should not be regarded as a complete or authoritative statement of law. The authority accepts no responsibility for the accuracy or otherwise for any interpretation or opinion in the law expressed herein. MoHUA also accepts no liability of any nature whether resulting from negligence or otherwise however caused arising from reliance upon the statements contained in this document.

Cities are expected to study this model policy and modify it as per their requirements. They may choose to forgo/add sections they deem fit for their own city.

Contents

Message from the Desk of Municipal Commissioner/IT HoD, <<City name>> Municipal Corporation

With the emergence of Smart Cities and data-driven innovation hubs across the country, the potential of data has become undisputed. Recognizing that timely and consistent access to data is an essential component of an open, transparent, collaborative and effective government, the city has decided to implement a City Data Policy (CDP) for the effective communication and coordination between citizens and government.

The main aim for designing this policy is to ensure data and information is utilised to its maximum potential. This policy will provide a framework for easier and effective data sharing among stakeholders and it is within the legal and legislative framework defined by the *<<City name>>* city. It is expected to trigger a higher quotient of trust towards governments, increase collaboration and engagement with citizens, public and private entities, lead to innovation-driven problem-solving, and eventually a higher quality of life for citizens. For this to happen, it is of paramount importance to have a set of guidelines with necessary conditions built in to avoid misuse, while also generating awareness among various stakeholders and the steps and processes that need to be put in place to enhance impact of data.

For the effective and timely implementation of this policy, I urge the city and all the stakeholders to treat this policy document as a call for participation in our joint effort towards creating a robust open governance and innovation ecosystem for a more liveable urban future.

Shri/Smt. *<<name>>*
Municipal Commissioner/IT HoD,
<<City name>> Municipal Corporation

I. Abbreviations/Definitions

- **Data:** Data refers to a representation of information, numerical compilations and observations, documents, facts, maps, images, charts, tables and figures, concepts in digital and/or analog form collected together for reference or analysis.

- **Data Archive:** A place where machine-readable data are acquired, manipulated, documented and distributed to others for further analysis and consumptions.

- **Data Generation:** Initial generation/collection of data or subsequent addition of data to the same specification.

- **Dataset:** A named collection of related sets of information composed of separate elements, but which can be manipulated as a unit.

- **Geospatial Data:** All data which is geographically referenced.

- **Information:** Processed data is referred to as Information.

- **Metadata:** Metadata is data about data. The information that describes the data source, and the time, place, and conditions under which the data were created. Metadata informs the users of who, when, what and where data were generated. Metadata allows the data to be traced to a known origin and known quality.

- **Negative List:** List of prohibitive datasets/feeds, deemed non-shareable by the departments/ organisations.

- **Restricted Data:** Data which are accessible only through a prescribed process of registrations and authorization by respective departments/organisation since it could lead to a threat to life or loss of public assets or critical infrastructure.

- **Shareable Data:** The data not covered under the scope of negative list and non-sensitive in nature falls under shareable data.

- **Standards:** Any application that embeds data handling functions (e.g. data collection, management, transfer, integration, publication etc.).

- **Open Access:** Access to data generated from public funding should be easy, timely, user-friendly and web-based without any process of registration/authorization.

II. Need for Data Policy

City Government departments generate a large amount of data. The data is generated as part of daily activities of the municipal corporations and city governance departments. This raw data, if used appropriately, can help generate valuable information for *<<City name>>* Municipal Corporation as well as entities external to *<<City name>>* Municipal Corporation. This data can be used by these various stakeholders for economic, scientific and developmental purposes.

The need to facilitate sharing and utilisation of this large amount of data primarily points to the need of a structure defining the rules and regulations. Today most of the government information is stored in databases, and is not effectively utilised for public good. The current regime of data management does not enable open sharing of Government-owned data with other arms of the Government, nor does it expect proactive disclosure of shareable data available with data owners. Such regimes could lead to duplication of efforts and loss of efficiency in planning activities focused on city development. Hence, City Data Policy (CDP) of *<<City name>>* Municipal Corporation aims to provide an enabling ecosystem and a platform for providing proactive and open access to the data generated through public investments and public revenue available with various departments of *<<City name>>* Municipal Corporation/*<<City name>>* Smart City Limited, other government departments etc. This document is intended to put in place a formal data governance mechanism at *<<City name>>* Municipal Corporation and is an attempt to further enhance the existing data initiatives of *<<City name>>* Municipal Corporation by having in place a robust, complete and inclusive CDP. Having such a policy in place will make the city of *<<City name>>* being identified as one of those cities that is data-driven and data-sufficient.

III. Scope of the Data Policy

This Policy document covers principle considerations concerning the use of data (access and utilisation of datasets including appropriate privacy management), and the principles governing the data sharing program for *<<City name>>*; thereby defining the expectations for departmental participation and governance of the data program.

The policy is intended as a resource for city administrators such as Municipal Commissioner, Smart City CEO, and other officials such as City Data Officer, heads of various government departments, Data Coordinators, Data Champions and external agencies - parastatal, civic, private) interested in engaging with the data initiatives of the City.

This Policy will apply to all data and information created, generated, collected and archived by *<<City name>>* Municipal Corporation/*<<City name>>* Smart City Limited. This policy applies to any person/user, organisation, administrators, contractors, etc. who intends to access information or assets through any data portal of *<<City name>>* Municipal Corporation/*<<City name>>* Smart City Limited. Specifically, the Data Policy applies to the following information assets of *<<City name>>* Municipal Corporation/*<<City name>>* Smart City Limited:

1. Data/information collected, captured, aggregated, processed and shared by *<<City name>>* Municipal Corporation/*<<City name>>* Smart City Limited
2. Citizens data/information
3. Personnel data/information relating to employees of *<<City name>>* Municipal Corporation/*<<City name>>* Smart City Limited

IV. Stakeholders and Collaboration

Following stakeholders will come together to set up City Data Alliance to assess, strategize, plan, implement and review the CDP:

a) **Government Agencies:** Government agencies operating with dedicated administrative structure in city namely Traffic Police, City Police, Central/State Government Departments, Government Autonomous Bodies etc. (apart from City Administration).

b) **Funding Agencies:** Funding agencies which regularly work with City Administration in different domains for e.g. World Bank, ADB, DFID, etc.

c) **Industry:** Key flagship manufacturing/service industry promoters/players in the city/state.

d) **Academia:** Representatives from leading universities/colleges/schools in the city.

e) **Policy Advocacy groups and NGOs:** Policy advocacy groups and NGOs working in different domains/areas like Slums, Health, Education, Environment, Participatory Governance, Mobility etc.

f) **Start-ups and Incubators:** Representatives from start-ups and incubators in the city/state.

g) **City Businesses:** Representatives from local small and medium business communities.

h) **Citizens and Communities:** Representatives from communities and citizen interest groups to further the interest of citizens/communities towards data driven policy governance and service delivery.

i) **Local Elected Representatives:** Local elected representatives to further the interest of citizens/communities towards data driven policy governance and policy formulation.

j) **Professional Representatives:** Representatives from various professional services like Doctors, CA, and Engineers etc.

Insert list/chart of stakeholders relevant to the Municipal Corporation for data initiatives

CITY DATA ALLIANCE (CDA)

The CDA will provide a collaborative framework to create and define use cases to solve critical city problems through the use of data, catalyse the right set of collaborations and networks to make available such data and undertake continuous dialogue between various stakeholders in the city around the CDP so as to inform and evolve the CDP effectively.

The responsibilities of the CDA will include:

a) To act as an advisory group to the city leadership on the CDP.

b) To assess the data needs of various smart city stakeholders.

c) To promote data driven governance and policy formulation.

d) To design and implement solutions based on city data.

e) To support industry to design solutions using emerging technologies like Artificial Intelligence (AI), Machine Learning (ML) and Blockchain.

f) To assess and design use cases critical to the citizens of the city.

g) To generate awareness among various stakeholders towards open government initiatives.

h) To bring city's stakeholders on a common platform to influence the city data priorities.

i) To facilitate data for co-creation and collaboration over civic issues.

j) To provide critical feedback to the city over the quality and relevance of data provided by the city.

k) To deliver research papers using city data on civic problems in the city.

l) To design and develop prototypes/ solutions on civic problems in the city.

m) To organise data-challenges on complex civic problems.

n) To organise hackathons and support shortlisted solutions at city level.

o) To set up scholarship for postgraduate and graduate students to work with the City Data Team along with the CDO.

p) To publish the progress report every month.

q) To prioritise the datasets/feeds for publishing on the open data platform.

r) To sensitise ecosystem partners to share data for solving civic challenges.

s) To support, engage and encourage network/groups/members of data enthusiasts in the city.

t) To improve city capacity over data driven governance and policy formulation.

u) To support CDOs by extending resources (like interns, researchers, technology experts), funds (program sponsorship etc.) and technology (solutions etc.).

v) To share data available with the partners on data platforms to promote city data.

V. Data Management and Team Structure

<<City name>> Municipal Corporation has set up a data team for management of data at the city level and enabling coordination between various departments for making sure that the data is available as and when required. The quality of data available and its use depends largely on the efficacy of the team that is put in place. The data team structure for <<City name>> Municipal Corporation is as follows:
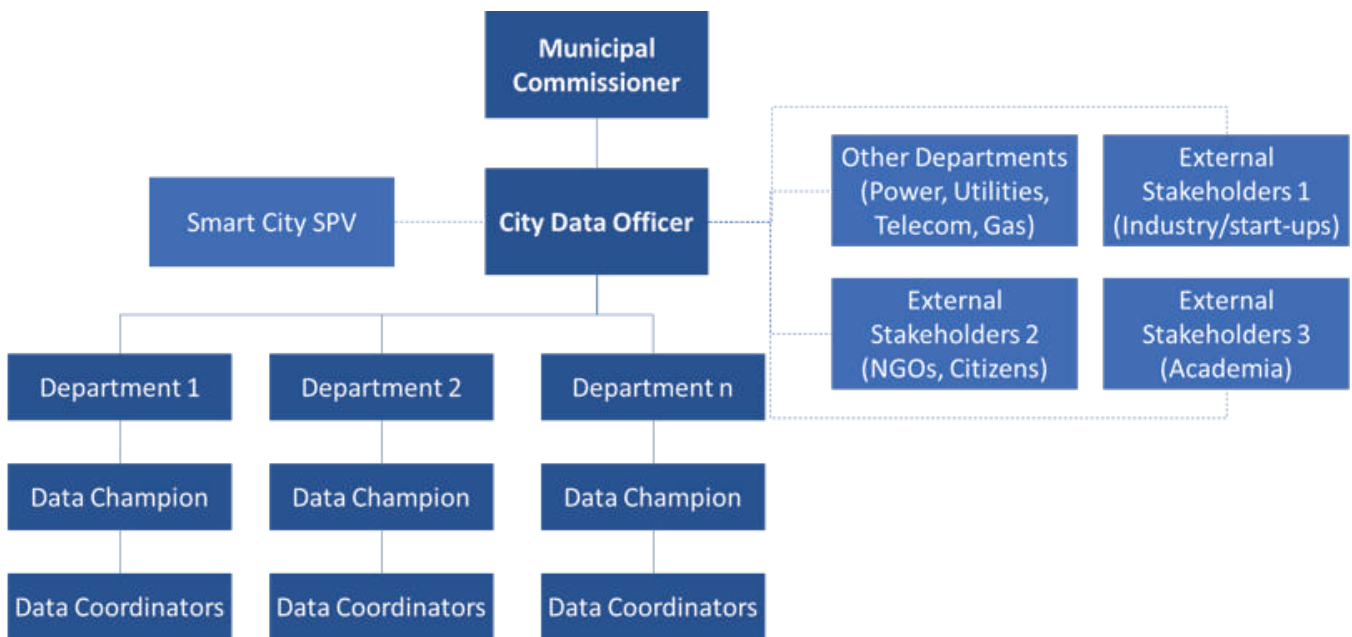
Insert specific data team structure chart, as per requirement

<<City name>> Municipal Corporation will also engage and secure buy-in from both internal and external stakeholders on key decisions. Further, this will help navigate through complicated hurdles (e.g. bureaucratic, political etc.) and to take prompt decisions and actions pertaining to collection, segregation and release of data.

VI. Roles and Responsibilities

Key officials of the data team and their roles and key responsibilities are defined below:

1. CITY DATA OFFICER (CDO)

The CDO will be the officer responsible for implementation of the DataSmart Cities Strategy at the city level. CDO will work with city leadership to assess and tap the potential of data and set up data culture across the organisation and outside the organisation. CDO will report directly to City Leadership and act as single point of contact to all internal and external stakeholders in the city. City leadership also needs to deploy dedicated skilled resources to drive the data initiative through CDO.

The key responsibilities of CDO are as follows:

a. Ensure that the CDP evolves as per the needs of various stakeholders of the city and relevant upgrades to policy are carried out time to time accordingly.

b. Coordinate with MDO (Mission Data Officer) to align with mission data strategy and priorities with respect to open government initiatives and policies.

c. Organise regular meetings of the City Data Alliance (CDA).

d. Coordinate with officers of various other government departments/agencies within the city for the effective implementation of CDP.

e. CDO along with team of Data Champions/ Coordinators must assess the data requirements of various stakeholders in smart city ecosystem. External stakeholders may also need to be engaged to understand the data needs. CDO must engage various internal stakeholder at operational, tactical and strategic level to assess the data need to make decisions. Data needs and frequency of consumption needs to be outlined for internal stakeholders.

f. Publish Data Catalogues and Datasets/Feeds on Open Government Data Portal and ensure that such datasets are updated at regular intervals as needed and create mechanisms for continuous feedback from citizens and stakeholders on type of datasets to be published.

g. Assess all the operational IT Projects for identifying public datasets/feeds. Data Champions and Data Coordinators in respective departments must prepare integration plan with respective IT vendor/ integrator to ensure compliance as per CDP.

h. Assess all proposed or under implementation projects to identify the datasets/feeds which could generate public datasets/feeds or may be useful for internal analysis. CDO must work with concerned System Integrator/vendor to ensure compliance of smart solutions with CDP.

i. Assess all periodic and recurring MIS needs to identify the datasets/feeds which could be shared with other departments through data exchange. CDO will also assess third party funded reports related to city operations for e.g. City Mobility plan, Health Plan etc.

2. DATA CHAMPIONS (DCs)

Data champions will be senior functionaries who would champion the implementation of the CDP in their respective departments/ organisations. Their responsibilities are as follows:

a. Shall identify the datasets/feeds, derived information, intelligence or data challenge with respect to day to day operations of the department.

b. Actively publish/enable publishing of datasets/feeds identified as relevant to the resolution of critical use cases for the city. They will work closely with the CDO for active implementation of the CDP.

c. DCs will be assisted by the Data Coordinators within the department to streamline processes of data reporting, collection and analysis etc. DCs will be responsible for data quality.

d. Undertake activities to engage with their stakeholders and evolve their department's strategy on data in line with the deliberations.

3. DATA COORDINATORS

Data Coordinators will assist Data Champions at the department/government agency level as reporting staff. Their responsibilities are as follows:

a. Aggregate the data demand from various channels.

b. Sensitizing the department employees over the importance of data quality etc.

c. Perform collection, interpretation and recording of data in accordance with CDP standards and CDO guidelines.

d. Perform data validation and ensure data quality.

e. Sort and organise the data; both hard copy and electronic versions.

f. Transmit data report to *<<City name>>* Municipal Corporation/*<<City name>>* Smart City Limited or CDO via Internet.

g. Update *<<City name>>* Municipal Corporation/*<<City name>>* Smart City Limited website or *<<City name>>* Open Data Portal with latest data records.

h. Assist department staff in data entry when required.

i. Provide data management updates in all internal and external meetings as required.

j. Analyse data for quality improvement purposes.

k. Prepare data for reporting, meetings and presentations for the concerned department and *<<City name>>* Municipal Corporation/*<<City name>>* Smart City Limited at large.

l. Ensure data management procedures comply with CDP.

m. Provide statistical analysis and longitudinal analysis of data.

n. Prepare and submit data required for audits.

VII. Data Management Principles

Implementing a data policy calls for a process that takes care of various aspects of data management. Having such a process in place will impart a sense of uniformity in the way in which data is generated/created to the point that it is destroyed/refined/reused. This process will be applicable to all types of data at all levels, all categories and classifications.

1. DATA CATEGORIZATION

Data will be categorized into two broad categories:

- **Personal Data:** Personal data is that data which is specific to a particular individual. It is the responsibility of the civic administration that they do not, in any case, publish personal identifiable data/information or parts of personal data/information on any of their Open Data Platforms or Datasets. It becomes the responsibility of the CDO to ensure that all personal data is anonymised before it is published.
- **Non-Personal Data:** Non-personal data is that data which cannot be identified or referenced to any individual. Anonymous data is also Non-personal data when all personal indicators and identifiers are eliminated for that particular data element.

2. DATA CLASSIFICATION

Data classification is the process of organising data into categories for its most effective and efficient use. Further to data categorization, there is a need to classify the data basis its intended usage and stakeholders. There may be some data which can be open to the public, while some other may be confidential and restricted. Such distinctions need to be appropriately defined to prevent misuse and maintain confidentiality. *<<City name>>* Municipal Corporation will prepare the negative list of data which will be periodically reviewed. Further, all datasets will be tagged such as open data, shareable, or others.

Personal and non-personal data will be broadly classified into four levels:
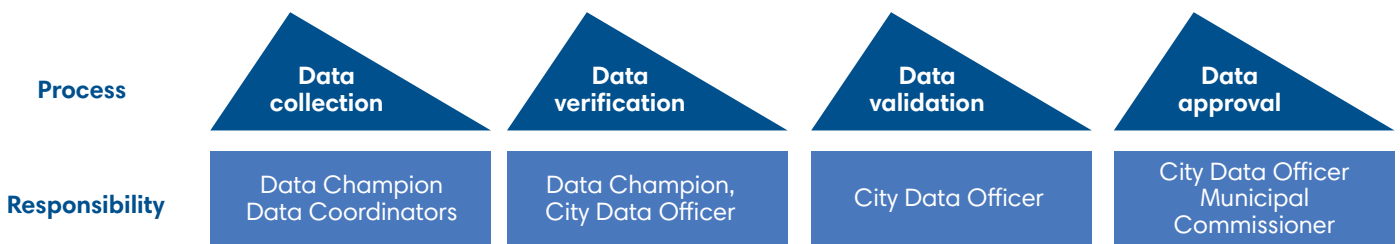
Classification	Class	Definition
Level 1	Public/Shareable Data	Those data not covered under the scope of negative list and non-sensitive in nature. This data is available for public consumption and use.
Level 2	Negative List	Non-shareable data as declared by the departments/organisations.
Level 3	Restricted Data	Data which are accessible only through a prescribed process of registration and authorization by respective departments/organisations.
Level 4	Sensitive Data	Sensitive data as defined in various Acts and rules of the Government of India.

3. DATA FLOW/APPROVAL FRAMEWORK

<<City name>> Municipal Corporation will set up enterprise processes to control the existing available data within the City administration. At every stage of data generation, the concerned stakeholders shall approve and authorise the data usage. Data flows may vary according to different scenarios, such as data being circulated between departments, uploading data on the open data portal, sharing data with third party, etc. Each dataset has a trustee accountable for data quality and security. Appropriate data flow and approval mechanisms should be in place for such situations. Also, all applications must be developed in compliance with the Government of India's India Enterprise Architecture (IndEA) framework.

The Approval framework for *<<City name>>* city data is as follows:

Process	Data collection	Data verification	Data validation	Data approval
Responsibility	Data Champion Data Coordinators	Data Champion, City Data Officer	City Data Officer	City Data Officer Municipal Commissioner

Please insert flowchart for approval framework, as per requirement

4. DATA ARCHIVAL AND RETENTION

Data Retention

CDO to undertake retention of data, i.e. the maintenance of documents to be accessed by an authorised user at a later stage. Retrieval schedule of the data will be as per the rules and regulations defined by the Government of India:

1) **Electronic:** Records to be retained in electronic form. These records may be stored on shared drives with access to only authorised individual/ group of individuals.

2) **Physical:** Record to be retained in physical form. These records may be kept in file cabinets or any other storage units assigned to each department with proper labelling so as to enable quick identification of the records.

3) **Original Form:** Records to be retained in the original form in which they were created or used i.e. either electronic or physical.

Data Archival

CDO to set-up the process for archiving data. Every dataset/feeds catalogue should contain archiving information. For specific file types (e.g. geo-spatial files), recent copy will be made available to users through Data Platform.

E-Files/records may be digitized by any one of the categories:

1) Category-I (e-Files/records to preserved permanently which are of historical importance) – For 10 years, it will be kept in the Department's server and thereafter transferred other available physical storage formats such as Tapes, hard-drives, Storages etc.

2) Category –II (e-Files/records of secondary importance and have a reference value for a limited period) – 10 years on the Department's server. In exceptional cases, if the record is required to be retained beyond 10 years it will be upgraded to Category-I.

Data will be stored in the main database for 6 Months in a live state so that whenever a report needs to be generated, the data will be extracted from main database. Data older than 6 months will be archived. If report duration extends beyond 6 months, the data will be retrieved from archival to generate the report.

Please change the time periods as appropriate and decided by the city authority

5. DATA SECURITY AND PRIVACY

CDO needs to ensure that data is protected from loss, unauthorised use and corruption, through adoption of international standards and best practices, duly protecting the privacy of personal data and confidentiality of sensitive data. All data flows, storage and sharing should adhere to the National and State level data privacy and security policies already in place. In case of any grievance, the associated parties may reach out to the CDO/Municipal Commissioner/Smart City CEO for resolution. Security measures needs to be followed during the collection and management of the classified data at all department levels, including:

• Minimize collection of personal data
• Delete data that is no longer necessary
• Restrict access to only those who need it
• Secure data throughout its entire lifecycle

The following security features will be followed while managing data in the city government:
• Data dissemination should be only to authenticated and authorised stakeholders (both internal and external) through data fiduciaries.
• The usage rules for data elements must specify for what purposes the data can or cannot be used. The solution should support SSL encryption mechanism for transferring data across network.
• The data transferred across network should be encrypted using Public Key Infrastructure (PKI).
• Access to all system resources including data files, devices, processes and audit files should be provided to the intended users only.
• All mobile applications should be designed and developed in a way that it ensures security of the application and data on the device.
• Ensure to protect documents by assigning security parameters and criteria in order to provide more effective protection for an electronic document in order to maintain Confidentiality, Authorization, Accountability, Integrity, Authenticity and Non-repudiation.
• Database Activity Monitoring (DAM) should be available to monitor all databases.

Ⅷ. **Standard Operating Procedures (SOPs)**

Standard Operating Procedures (SOPs) are succinct guidelines designed to achieve consistency in specified situations by postulating a standard practice in performing those functions. These SOPs are designed with a view to enhance and standardise data operation and management.

6.a) SOP FOR DATA COLLECTION

Data collection is referred to as the method of collecting information in a systematic way. This is the first step towards data-driven decision making and evidence-based governance. Data collection provides both a baseline to measure and a target to improve. In order to get the primary datasets we have to collect data from different wards, zones which is under *<<City name>>* Municipal Corporation, *<<City name>>* Smart City Limited employee and Other Government departments. Data which is not going to be used for any kind of analysis or will not be used for any communication purpose should not be collected at all.

If the request is received from external agency:

1. If the request is received from external agency, it should be directed to the CDO.
2. Depending on the requested data or the data which need to be collected, the CDO shall direct the request to Data Champion (DC) of the concerned department.
3. The DC of the concerned departments checks for the requested data. If data is available with department, DC shall instruct the Data Coordinators or the concerned personnel to gather the data in requested format.
4. DC will take approval of data from their HOD.
5. If fresh data is needed to be captured/acquired, the Data Champion/Data Coordinators in consultation with CDO shall take appropriate action.

If the request is received from internal departments:

1. If the request is received from internal departments, it should be directed to the Data Champion for the concerned department.
2. The DC of the concerned departments checks for the requested data. If data is available with department, DC shall instruct the Data Coordinators or the concerned personnel to gather the data in requested format.
3. If fresh data is needed to be captured/acquired, the Data Champion/Data Coordinators in consultation with CDO shall take appropriate action.

6.b) SOP FOR ELECTRONIC DATA COLLECTION

Additional factors need to be kept in mind when any form of data is collected, accessed, transferred or stored electronically. Cities ecosystem comprises of various technology solutions ranging from Sensors, IoT, SCADA, Electronic camera, GIS, payments system etc. which generate loads of structured data every second on different dimensions. Cities could also leverage various unstructured data emerging from different sources and third-party systems like social media, internet, websites, videos, images etc.

1. The data should be collected with consent of the end-user who may be a citizen or *<<City name>>* Municipal Corporation/*<<City name>>* Smart City Limited employee.

2. Special care needs to be taken for data privacy and security. The measures mentioned in above sections of the Policy shall be followed to maintain confidentiality and security of data.

3. For data collection and integration with the IoT systems / sensors, the following Data Architecture highlighted below shall be followed by *<<City name>>* Municipal Corporation/*<<City name>>* Smart City Limited.

Insert City Specific Data architecture/flow
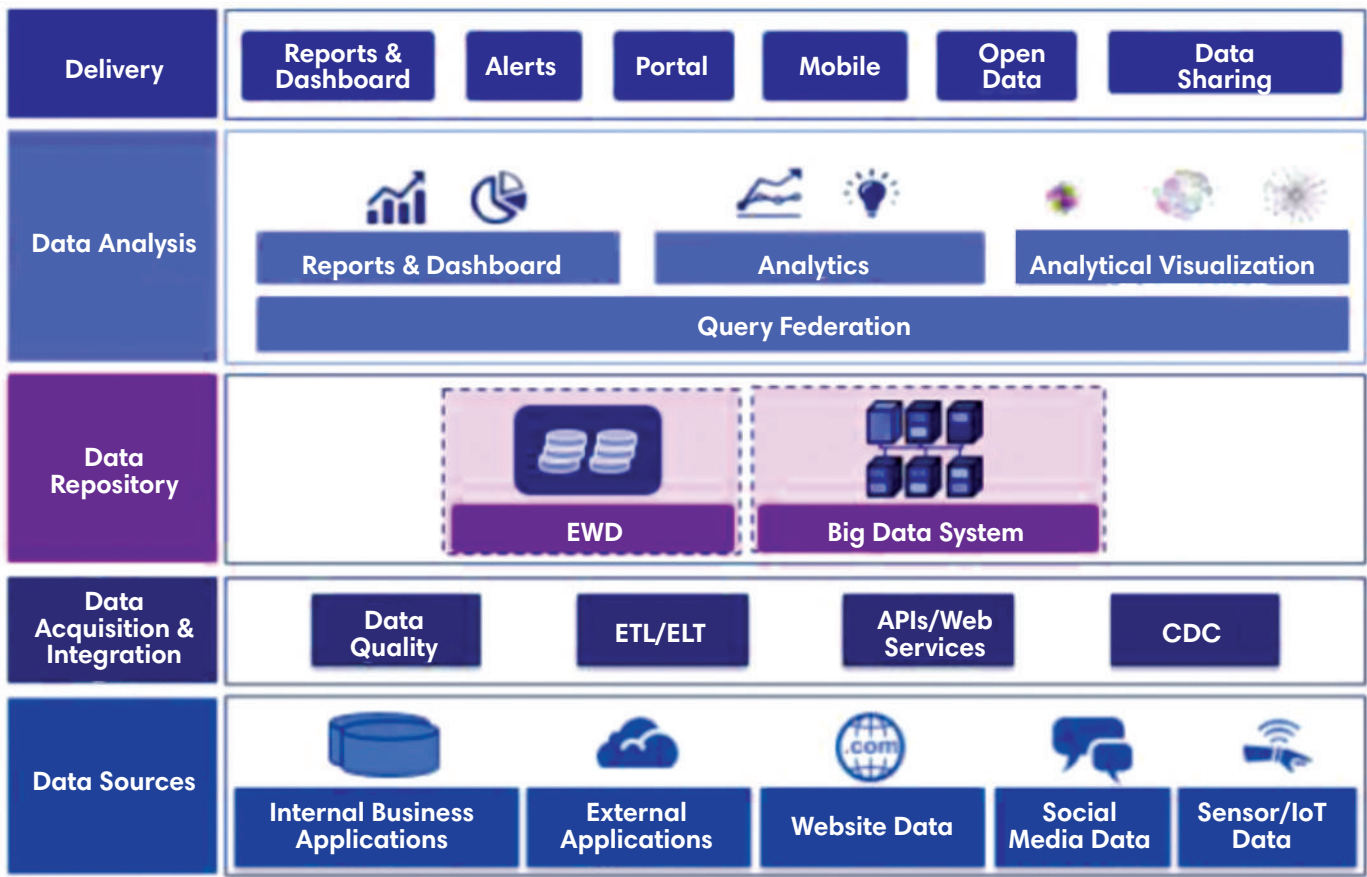
7. SOP FOR DATA PROCESSING AND CLEANING

Raw data may be old and inaccurate and can have an adverse impact on results. Data cleaning will be done to ensure that data is correct, consistent and useable by identifying any errors or corruptions in the data, correcting or deleting them, or manually processing them as needed.

1. While collecting the electronic data, the IT applications/ IT systems should be developed in such a way that under any circumstances these applications/systems should not accept any wrong data/null data.

2. If there is existing data, identify discrepancies which may come from different sources.

3. The collected data shall be properly processed and cleaned before performing any kind of analysis.

4. If needed, commercial software available in the market can be used with prior approvals from the concerned authorities.

Extract, Transform and Load (ETL) is the common methodology used for data integration and processing. It is a three-step process which used for data integration to blend data from multiple sources. It's often used to build a data warehouse. During this process, data is taken (extracted) from a source system, converted (transformed) into a format that can be analysed, and stored (loaded) into a data warehouse or other system.

Specifically, Data Champion/ Data Coordinator will keep in mind the following points while collecting data from their respective department:

- Spaces in extra columns Compliance
- Ward-wise Compliance
- Blank Cells Compliance
- Standard format Compliance
- All NA Compliance
- Special Characters Compliance
- Split Sheets Compliance
- Datasets Compliance
- Data Completeness

8. SOP FOR QUALITY ASSESSMENT OF DATASETS

Quality assessment of data is needed to ensure that the quality standard is met i.e. accuracy, free from any sort of legal issues, privacy of an individual is maintained and does not compromise with the National security. Data quality shall be assessed from the perspectives of adequacy, appropriateness, accuracy and reliability, authenticity, consistency and validity.

- Responsibility for quality check of data rests with the CDO while publishing data on the open data portal or sharing it with stakeholders upon request.
- Under some circumstances, a special committee under the leadership of Municipal Commissioner (*<<City name>>* Municipal Corporation), comprising of CDO, and Additional/Assistant Commissioner may be formed for data quality assessment.

9. SOP FOR DATA PUBLISHING

National Data Sharing and Access Policy (NDSAP) defines standards for publishing datasets and feeds on the open data portal. CDOs must ensure adherence towards defined standards and classification. This SOP describes the steps required for publishing data to the Smart Cities Open Data Portal:

1. **Understand the requirement:** Follow proper procedures to collect the relevant data to be uploaded on the open data portal. Understand the publishing options and the available datasets.
2. **Process the data:** Ensure data is in an appropriate format to be published on the open data portal. It does not contain any personalized information, is open, authenticated and free from defects.
3. **Prepare to deploy/publish data:** Follow procedures specific to the publishing option you have selected and work with the appropriate team (when necessary) to publish your data. NDSAP recommends that datasets should be published in an open format and should be machine readable. Data format can be chosen from the list recommended by NDSAP, highlighted in the Appendix.
4. **Publish metadata:** Follow established metadata procedures as per NDSAP and any other guidelines laid down by the city to publish metadata on the Portal and create linkages between data and metadata.
5. **Obtain approvals and finalise deployment:** Obtain the appropriate management approvals for your data based on your selected data publishing option, either from CDO or Municipal Commissioner.

10. SOP FOR ENGAGING STAKEHOLDERS

The concept recognizes the value of enhancing engagement among all four stakeholders of the quadruple-helix model—Government (*<<City name>>* Municipal Corporation Departments, *<<City name>>* Smart

City Limited and Other Departments of Governments), citizens, academia, and industry, along with improvements in the internal workflow and decision-making processes of city governments.

CDOs along with the team of Data Champions/Coordinators shall assess and document the data requirements of various stakeholders in the city ecosystem, along with frequency of consumption and level of granularity.

Key activities may include:
- Identifying stakeholders from various age groups and ethnicities and engaging them in city initiatives
- Organising workshops, hackathons/events to promote brainstorming over required datasets
- Decision making and consultation with data experts to zero down on the most important datasets required on the portal
- Data ideation with public forum to gain understanding of citizen/industry needs
- Formation of city data alliances

11. SOP FOR DATA COLLECTION, PROCESSING AND ANALYSIS FOR ON-FIELD SURVEYS

Field survey is defined as collection and gathering of information at the local level by conducting primary surveys. On-field surveys may be required in situation where data from digital sensors or existing datasets are inadequate. Surveys may be administered to fulfil a certain gap, with a well-defined problem for investigation. Data collected from on-field surveys can help *<<City name>>* Municipal Corporation/*<<City name>>* Smart City Limited get a snapshot of how things are at a specific time. The survey research may be descriptive, analytical or evaluative. Field surveys are cost-intensive hence they may be conducted after a thorough mapping of their objectives and outcomes.

Survey data processing consists of four important steps:
1. **Survey designing:** Survey is a research strategy and not a method. CDOs/concerned team should choose the most appropriate method based on purpose. CDO shall help the concerned team devise suitable surveys/questionnaires by clarifying the objective, determining sample and deciding upon the survey mode to finally create the questionnaire. A research method should not advise the questions, but other way around. Posing difficult to answer questions, in such case a simple rule or scale can be used to help respondents.

2. **Data collection:** Survey can employ a range of methods including questionnaires, interviews or even focus group discussions. Data entry in either format may happen manually or electronically. Going-forward, all the e-governance IT applications/systems shall be designed in such a way that manual processes get replaced by automated process without much intervention of humans. As most of process would be automated and handled by an e-mode, data will be available for further analysis.

3. **Data processing:** Before any analysis is possible, ensuring accuracy and quality of data is paramount. Survey form data is always prone to errors, omissions and other inconsistencies. This data inconsistency and incompleteness, if not edited and corrected on time, can complicate the analysis and may even result in wrong analysis. Data processing shall comprise of various steps necessary for preparing the data for analysis, including editing, data classification, removing redundancies, and preparation of

tables. This is an important step when the survey instrument collects qualitative data, which needs to be then represented in a format for analysis.

4. **Data analysis:** Data analysis covers the final step of characterizing and interpreting research findings. In situations where the digital tools are employed for the survey, and the data can be processed easily. Data analysis will involve computation of certain indices or measures along with searching for patterns of relationship that exist among the data groups. The task of analysing quantitative data may be accomplished through statistics. Descriptive statistics is to be used for organising raw data obtained in the process of research, such as tabulation and classification of data. Inferential statistics, also known as sampling statistics, will be used for making inferences or conclusions from the data collected from a sample and drawing generalisations on the entire population.

12. SOP FOR DATA ANALYSIS

Data analysis or analytics is an encompassing and multi-dimensional field that uses mathematics, statistics, predictive modelling and machine-learning techniques to find meaningful patterns and knowledge in recorded data. This will help in using intelligent techniques to uncover actionable insights from the relevant data.

<<City name>> Municipal Corporation/*<<City name>>* Smart City Limited to establish analytical capabilities within the Data Team to accomplish data analysis on the cleaned and processed data. Various examples not narrowing down to the same can be sited as follows:
a) Slice and dice to drill down the data till lowest entity
b) Trend analysis and pattern identification on time series (days, weeks, months, quarter or seasonal, etc.),
c) Trend analysis and pattern identification using various dimensions: Cost, Budget, domain specific parameters etc.
d) Comparison between various parameters in different geographies etc.
e) Visualization to view the trends and patterns for decision making. Converting the data into a more comprehensible and user-friendly format.

Basis the data collected, the Data Team of *<<City name>>* Municipal Corporation/*<<City name>>* Smart City Limited under the leadership of CDO shall perform various analytical tasks as under:
• Descriptive Analytics: It helps in answering "What is happening?"
• Diagnostic Analytics: It helps in answering "Why did it happen?"
• Predictive Analytics: It helps in answering "What is likely to happen?"
• Prescriptive Analytics: It helps in answering "What should I do about it?"

IX. Annexure

1. OPEN GOVERNMENT DATA (OGD) PLATFORM POLICY

Government of India has listed down policy for external users to access the data hosted on the Open Government Data Platform. This policy can be access at **https://data.gov.in/policies**. The same needs to be implemented and followed by *<<City name>>* while implementing the policy document. Further, the Government Open Data Use Licence – India has been approved. This can be accessed at **https://data.gov.in/sites/default/files/Government_Open_Data_Use_Licence_India.pdf**

2. STANDARDIZATION OF DATA ACCESS AS PER NATIONAL DATA SHARING AND ACCESS POLICY (NDSAP)

NDSAP defines standards for publishing datasets and feeds. These standards need to be adhered to by *<<City name>>* while sharing its datasets.

Open Source Driven: Datasets are considered to be open by default unless classified as internal, sensitive, protected or restricted.

Metadata: Datasets and apps must be published along with proper metadata. Besides facilitating easy access to datasets, using a common data taxonomy/structure shall be extremely useful in the future for federation/integration of data catalogues. Key Metadata elements for catalogues/resources include:

CATALOGUES

- **Title (Required):** A unique name for the catalogue (group of resources) viz. Current Population Survey <Year>, Consumer Price Index <Year>, Variety-wise Daily Market Prices Data, City-wise Construction of Deep Tube wells over the years, etc.
- **Description (Required):** Provide a detailed description of the catalogue e.g., an abstract determining the nature and purpose of the catalogue.
- **Keywords (Required):** It is a list of terms, separated by commas, describing and indicating at the content of the catalogue. Example: rainfall, weather, monthly statistics.
- **Group Name (Optional):** This is an optional field to provide a Group Name to multiple catalogues in order to show that they may be presented as a group or a set.
- **Sector & Sub-Sector (Required):** Choose the sectors(s)/sub-sector(s) those most closely apply to your catalogue.
- **Asset Jurisdiction (Required):** This is a required field to identify the exact location or area to which the catalogue and resources (dataset/apps) caters to viz. entire country, state/province, district, city, etc.

RESOURCES (DATASETS/APPS)

- **Category (Required):** Choose from the drop-down options of whether it is a Dataset or an Application.
- **Title (Required):** A unique name of the resource viz. Consumer Price Index for <Month/Year> etc.
- **Access Method (Required):** This could be "Upload a Dataset" or "Single Click Link to Dataset".

- **Reference URLs:** This may include description to the study design, instrumentation, implementation, limitations, and appropriate use of the dataset or tool. In the case of multiple documents or URLs, please delimit with commas or enter in separate lines.
- **Access Type:** It mentions the type of access viz. Open, Priced, Registered Access or Restricted Access (G2G).
- **Date Released:** It mentions the release date of the Dataset/App.
- **Note:** It mentions the anymore information the contributor/controller wishes to provide to the data consumer or about the resource.
- **NDSAP Policy Compliance:** This field is to indicate if this dataset is in conformity with the National Data Sharing and Access Policy of the Govt. of India.

IF RESOURCE CATEGORY IS DATASET
- **Frequency (Required):** It mentions the time interval over which the dataset is published on the OGD Platform on a regular interval (one-time, annual, hourly, etc.).
- **Granularity of Data:** It mentions the time interval over which the data inside the dataset is collected/updated on a regular basis (one-time, annual, hourly, etc.).

IF RESOURCE CATEGORY IS APP
- **App Type (Required):** It mentions the type of App being contributed viz. Web App, Web Service, Mobile App, Web Map Service, RSS, APIs etc.
- **Datasets Used:** Datasets used for making this app.
- **Language:** Language used for app.

Data Formats: NDSAP recommends that datasets should be published in an open format, which can be accessed without the need for a software licence and should be machine readable. The data could be published in any of the following formats:

- CSV (Comma separated values)
- XLS (Spreadsheet - Excel)
- ODS (Open Document Formats for Spreadsheets)
- XML (Extensive Mark-up Language)
- RDF (Resources Description Framework)
- KML (Keyhole Mark-up Language used for Maps)
- GML (Geography Mark-up Language)
- RSS/ATOM (Fast changing data e.g. hourly/daily)

NDSAP can be accessed at **https://data.gov.in/sites/default/files/NDSAP.pdf.** Guidelines for implementation of NDSAP are available at **https://data.gov.in/sites/default/files/NDSAP%20 Implementation%20Guidelines%202.4.pdf**

3. STATE LEVEL DATA POLICY
The *<<State name>>* Data Policy is designed so as to apply to all shareable non-sensitive data available either in digital or analog forms and generated using public funds by various State departments and Subordinate offices/organisations/agencies. It is designed to promote data sharing and enable access to Government-owned data that could be used for planning and development.